

Голові разової спеціалізованої
вченої ради PhD 8729
Хмельницького національного університету
доктору технічних наук, професору
Тетяні ГОВОРУЩЕНКО

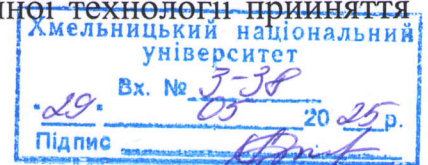
Рецензія
на дисертаційну роботу Молчанової Марини Олексіївни
на тему «Методи виявлення та класифікації прийомів та об'єктів
пропаганди у текстовому контенті засобами штучного інтелекту»,
що подана на здобуття наукового ступеня доктора філософії
в галузі знань 12 Інформаційні технології
за спеціальністю 122 Комп'ютерні науки

Актуальність теми та зв'язок з науковими планами наукових робіт університету. В умовах активного використання інформаційних технологій як інструменту впливу на суспільну свідомість, виникає потреба у розробці методів, здатних ефективно ідентифікувати як об'єкти пропаганди, так і специфічні мовні та стилістичні прийоми, що використовуються для її реалізації.

Запропоноване в дисертації поєднання методів глибокого навчання з технологіями обробки природної мови відповідає сучасному напрямку розвитку комп'ютерних наук та штучного інтелекту. Розробка методів до виявлення структурних ознак пропагандистських текстів із можливістю їх класифікації сприяє створенню практично придатних інструментів для виявлення небажаного або маніпулятивного контенту в інформаційному просторі. Запропонований підхід передбачає комплексний аналіз текстів з урахуванням семантичних і структурних характеристик, що забезпечує можливість встановлення взаємозв'язків між використовуваними пропагандистськими прийомами та їх об'єктами.

Таким чином, тематика дослідження узгоджується з сучасними науковими підходами у галузі комп'ютерних наук, є науково-прикладною за своєю спрямованістю та орієнтована на вирішення актуальних завдань інформаційної безпеки.

Дослідження, результати яких викладено в дисертаційній роботі, проведено в процесі виконання держбюджетної теми Хмельницького національного університету «Розроблення інформаційної технології прийняття



контрольованих людиною критично-безпекових рішень за ментально-формальними моделями машинного навчання» (ДР № 0121U112025), де авторка була виконавицею окремих задач.

Формулювання наукової задачі, мети й задач дослідження. У результаті виконання дисертаційної роботи здобувачкою було розв'язано актуальну науково-прикладну задачу автоматизованого виявлення прийомів та об'єктів пропаганди, що дозволяє комплексно аналізувати взаємозв'язки виявлених прийомів і об'єктів. Було правильно визначено об'єкт і предмет дослідження, відповідно до наведеної гіпотези дослідження. Об'єкт дослідження формалізовано як процес інтелектуального аналізу текстового контенту для виявлення прийомів та об'єктів пропаганди. Предметом дослідження встановлено методи та засоби обробки природної мови для виявлення прийомів та об'єктів пропаганди.

Мету дисертаційної роботи визначено, як є підвищення точності та якості виявлення прийомів та об'єктів пропаганди за семантичними маркерами у текстовому контенті засобами штучного інтелекту з подальшим поясненням прийнятих рішень.

Поставлену мету роботи досягнуто в результаті розв'язання таких задач:

- 1) проведено аналіз методів, засобів та технологій для автоматизованого виявлення пропаганди у текстовому контенті;
- 2) удосконалено метод класифікації текстів за вмістом пропаганди нейромережевими моделями глибокого навчання;
- 3) розроблено метод виявлення прийомів пропаганди за маркерами із візуальною інтерпретацією прийнятих рішень;
- 4) розроблено метод виявлення об'єктів пропаганди нейромережевими моделями глибокого навчання з візуальною інтерпретацією прийнятих рішень;
- 5) розроблено інтелектуальну інформаційну систему для валідації запропонованих методів і проведено експериментальні дослідження.

Формулювання задач є коректним і повністю узгодженим із загальною структурою роботи.

Наукова новизна одержаних авторкою результатів полягає у наступному:

- 1) удосконалено метод класифікації текстів за вмістом пропаганди нейромережевими моделями глибокого навчання, який відрізняється від аналогів модифікованою архітектурою нейромережі та обсягом вихідних даних, що дало змогу підвищити точність класифікації.

2) розроблено новий метод виявлення прийомів пропаганди за маркерами з візуальною інтерпретацією прийнятих рішень, який відрізняється від аналогів використанням доповненої множини семантичних маркерів для виявлення прийомів пропаганди, що дало змогу пояснити отримані результати і підвищити точність та якість виявлення пропаганди.

3) розроблено новий метод виявлення об'єктів пропаганди нейромережевими моделями глибокого навчання з візуальною інтерпретацією прийнятих рішень, який відрізняється від аналогів асоціативним групуванням об'єктів пропаганди, що дало змогу покращити результати виявлення об'єктів пропаганди та візуально їх інтерпретувати.

Обґрунтованість і достовірність наукових положень, висновків і рекомендацій. Наукові положення, сформульовані в дисертаційній роботі, мають належне теоретичне обґрунтування та підтверджуються результатами експериментальних досліджень. Запропоновані методи базуються на сучасному апараті обробки природної мови, машинного навчання та нейромережових моделей, що забезпечує методологічну коректність проведених розрахунків. Експериментальна частина містить достатній обсяг емпіричних даних, що використовуються для верифікації результатів. Вибір метрик, порівняння з базовими та актуальними підходами, а також аналіз отриманих результатів свідчать про належний рівень обґрунтованості зроблених висновків. Рекомендації, сформульовані в роботі, ґрунтуються на результатах проведених досліджень та можуть бути використані при побудові систем виявлення інформаційного впливу в текстовому контенті, що підтверджує практичну значущість отриманих результатів. Результати дисертаційної роботи мають практичне впровадження на підприємствах, що демонструє відповідність теоретичних досліджень з реальними результатами застосування.

Практичне значення одержаних результатів. Розроблена інтелектуальна система комплексного виявлення та класифікації прийомів та об'єктів пропаганди у текстовому контенті засобами штучного інтелекту, що надає користувачам можливість автоматизованого аналізу текстів, забезпечуючи комплексний результат у вигляді: загальної оцінки прояву пропаганди у тексті; виявлених пропагандистських прийомів із числовими характеристиками сили їх прояву; оцінки приналежності виявлених об'єктів пропаганди до використаних прийомів. Окрему увагу приділено модулю візуалізації, який надає користувачеві інтерпретовані результати аналізу, що підвищує прозорість функціонування системи та сприяє формуванню довіри до отриманих висновків.

Результати експериментального дослідження підтверджують ефективність запропонованих методів підвищення точності виявлення пропаганди в текстовому контенті. Зокрема, використання нейромережових моделей глибокого навчання для класифікації текстів за наявністю пропаганди дозволяє досягти приросту точності на 0,035 за метрикою Ассигасу і щонайменше на 0,06 за метрикою F_1 у порівнянні з існуючими підходами. Запропонований підхід до виявлення пропагандистських прийомів на основі маркерів з візуальною інтерпретацією рішень забезпечує підвищення точності на 0,368 (Ассигасу) порівняно з відомими аналогами. Застосування візуальної аналітики (зокрема LIME) у структурі системи сприяє підвищенню пояснюваності результатів, зокрема шляхом порівняння отриманих значень із еталонними маркерами, а також дозволяє здійснювати інтерпретацію виявлених об'єктів пропаганди, групування виявлених об'єктів і визначення їх приналежності до відповідних прийомів. Сукупність зазначених рішень забезпечує підвищення достовірності та якості ідентифікації проявів пропаганди у текстовому контенті.

Теоретичні та практичні результати дослідження впроваджено у: відділі протидії кіберзлочинам в Хмельницькій області Департаменту кіберполіції Національної поліції України (довідка про впровадження від 19.03.2025 р.); ГО «ІТ Кластер м. Хмельницького» (довідка про впровадження від 10.12.2024 р.); ПП «Авіві» (довідка про впровадження від 25.10.2024 р.); ТОВ «Системи для бізнесу 2» (довідка про впровадження від 17.01.2025 р.); навчальному процесі Хмельницького національного університету (акт впровадження від 27.11.2024 р.).

Особистий внесок здобувача полягає в розробленні методів та відповідної інтелектуальної системи, що забезпечує розв'язання поставлених у дисертації задач. Усі основні наукові та прикладні результати дисертаційної роботи отримані здобувачкою самостійно. За результатами проведених наукових досліджень основні наукові результати опубліковано в 12 наукових працях: 4 статті у фахових наукових журналах України, що включені на дату опублікування до переліку наукових фахових видань України категорії Б; 5 публікацій, що засвідчують апробацію матеріалів дисертації (індексуються в наукометричній базі Scopus); 3 авторських свідоцтва. У роботах, що опубліковані в співавторстві, авторці належать основні ідеї, теоретична та практична розробка положень, що відображені в характеристиці наукової новизни, а саме метод класифікації текстів за вмістом пропаганди нейромережевими моделями глибокого навчання; метод виявлення прийомів

пропаганди за маркерами із візуальною інтерпретацією прийнятих рішень; метод виявлення об'єктів пропаганди нейромережевими моделями глибокого навчання з візуальною інтерпретацією прийнятих рішень.

Окрім того, положення дисертаційної роботи апробовано на п'яти міжнародних наукових конференціях, за матеріалами яких підготовлено п'ять публікацій, що індексуються в наукометричній базі даних Scopus. У кожній публікації, виконаній у співавторстві, чітко визначено особистий внесок здобувача, що свідчить про відповідальне дотримання вимог до апробації наукових результатів.

Апробація матеріалів дисертації. Основні положення та наукові результати доповідались на міжнародних і всеукраїнських науково-технічних та науково-практичних конференціях і семінарах, а саме: 12th International Conference on Advanced Computer Information Technologies «ACIT'2022» (September 26–28, 2022), 14th International Scientific and Practical Programming Conference «UkrPROG 2024» (May 14–15, 2024, Kyiv, Ukraine), 12th International Conference Information Control Systems & Technologies «ICST 2024» (September 23–25, 2024, Odesa, Ukraine), Information Technology and Implementation Workshop: Intelligent Systems and Security «IT&I-WS 2024: ISS» (November 20–21, 2024, Kyiv, Ukraine), 7th Workshop for Young Scientists in Computer Science & Software Engineering «CS&SE@SW 2024» (December 27, 2024, Kryvyi Rih, Ukraine).

Структура та обсяг дисертації. Дисертаційна робота складається з анотації, змісту, переліку умовних скорочень, вступу, чотирьох розділів, висновків, списку використаних джерел із 156 найменувань на 23 сторінках і 4 додатків. Загальний обсяг дисертаційної роботи становить 179 сторінок друкованого тексту, із них 138 сторінок основного тексту. Дисертація містить 43 рисунки і 15 таблиць.

Зауваження. У результаті аналізу рукопису було виділено такі зауваження:

1) У пункті 3.2 (або під час розгляду, в пункті 1.6) варто більш детально навести критерії відбору семантичних маркерів, використаних для виявлення пропагандистських прийомів.

2) У роботі не враховано аспект історичної тривалості та соціокультурного контексту текстів. Зокрема, незрозуміло, як система реагує на тексти, написані багато десятиліть тому (наприклад, класична література), де могли бути закладені приховані форми впливу.

3) В рамках методу виявлення прийомів пропаганди за маркерами із візуальною інтерпретацією прийнятих рішень (п. 3.2) застосовано метод LIME для візуальної інтерпретації рішень моделі, однак не подано порівняння з іншими підходами пояснення рішень нейронних мереж (наприклад, SHAP, Integrated Gradients тощо), що дозволило б обґрунтувати доцільність саме такого вибору. За відсутності порівняльного аналізу доцільно навести аргументи щодо вибору LIME як базового інструменту інтерпретації, з урахуванням особливостей поставленої задачі та специфіки вхідних даних.

4) Розгляд пунктів 1.2 та 1.3 подано фрагментарно, без впорядкування висновків. Відсутня порівняльна таблиця, яка б систематизувала існуючі підходи та їхні характеристики.

5) Необхідно чіткіше розмежувати пропаганду від суміжних понять, таких як фейкова інформація, маніпуляція, дезінформація. На с. 16–20 (п. 1.2) опис даних категорій частково перетинається, але без систематизації та уточнення методологічних розбіжностей.

6) Наведений в пункті 3.2 метод виявлення прийомів пропаганди за маркерами із візуальною інтерпретацією прийнятих рішень було б доцільним подати як два окремих методи – метод виявлення маркерів пропаганди та метод виявлення прийомів пропаганди.

7) У пункті 4.3 відсутній аналіз помилок моделей (false positives/negatives), що є важливим для задач соціально-значущої класифікації.

Втім, зазначені зауваження істотно не впливають на зміст дисертаційної роботи та не знижують її наукову новизну та практичну цінність.

Загальний висновок. З огляду на вищесказане, вважаю, що дисертаційна робота Молчанової М.О. «Методи виявлення та класифікації прийомів та об'єктів пропаганди у текстовому контенті засобами штучного інтелекту» містить нові науково-обґрунтовані теоретичні та експериментальні результати в предметній області спеціальності 122 «Комп'ютерні науки» в галузі 12 «Інформаційні технології», які в сукупності забезпечують розв'язання актуальної науково-прикладної задачі виявлення та класифікації пропагандистських прийомів і об'єктів у текстовому контенті, яка дозволяє ефективно виявляти та класифікувати прийоми пропаганди та аналізувати інформаційні загрози для використання у медіа-ресурсах та соціальних мережах, що є важливим кроком у боротьбі з дезінформацією та пропагандою.

Отже, дисертаційна робота «Методи виявлення та класифікації прийомів та об'єктів пропаганди у текстовому контенті засобами штучного інтелекту», яка подана на здобуття ступеня доктора філософії, за своїм науковим рівнем та

практичною цінністю, змістом та оформленням відповідає вимогам пп. 6, 7, 8, 9 «Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії», затвердженого Постановою Кабінету Міністрів України від 12 січня 2022 р. №44 (зі змінами, внесеними згідно з Постановами Кабінету Міністрів України № 341 від 21.03.2022, № 502 від 19.05.2023, № 507 від 03.05.2024), а її авторка, Молчанова Марина Олексіївна, заслуговує на присудження ступеня доктора філософії за спеціальністю 122 «Комп'ютерні науки».

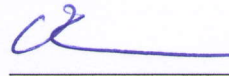
Рецензент

доктор технічних наук, професор

кафедри комп'ютерної інженерії

та інформаційних систем

Хмельницького національного університету

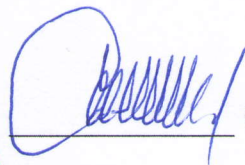
 Олег САВЕНКО

«Підпис Олега САВЕНКА засвідчую»:

Проректор з наукової роботи

Хмельницького національного університету



 Олег СИНЮК